

**Metodología para crear un perfil de cliente productor de caña panelera para otorgamiento de crédito a partir de variables continuas**

Amy Jessica Abadiano Zuleta.  
Noviembre 2019.

Fundación Universitaria Los Libertadores.  
Especialista en Economía Aplicada.

Metodología para crear un perfil de cliente productor de caña panelera para otorgamiento de crédito a partir de variables continuas

---

Presentado por

**Amy Jessica Abadiano Zuleta**

en cumplimiento parcial de los requerimientos para optar al título

LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

de

**Especialista en Estadística Aplicada**

*Dirigida por*

**Patricia Gallego Torres**



---

---

---

---

---

# LOS LIBERTADORES

## FUNDACIÓN UNIVERSITARIA

---

Firma del presidente Del jurado

---

Firma del jurado

---

Firma del jurado

### **Agradecimientos**

El desarrollo del trabajo fue posible gracias a la profesora Patricia Gallego quien con su experiencia, sugerencias y recomendaciones se culminó el desarrollo del trabajo a la Universidad y al coordinador Sebastián por brindarme el apoyo y facilidades en mi aprendizaje

## **Dedicatoria**

El presente trabajo de investigación lo dedico a las personas que más quiero en el mundo a mi esposo por su Fortaleza, comprensión y amor incondicional, a mi mami por su cariño, ayuda y por enseñarme a ser fuerte en las adversidades, a mi hijo por su amor, por ser el motor de mi vida y mis alegrías y a Dios porque me ha dado la salud y la perseverancia para seguir adelante y cumplir cada meta y sueños propuestos.

El presente trabajo tiene como objetivo segmentar los clientes aptos para otorgamiento de crédito en caña panelera a través de los indicadores financieros presentados en el los estados financieros y balance, se vio la necesidad de realizar este estudio ya que muchas entidades financieras evalúan demasiados indicadores al momento de otorgar un préstamo y esto incurre en demora en los tiempos de respuesta para los clientes, otros bancos tienen tiempos de respuesta mas rápidos y ágiles en el proceso de desembolso la idea es estar a la vanguardia del Mercado y poder ser competitivos, por esta razón se usó la metodología de análisis de componentes principales para reducción de variables financieras y una vez obtenido los factores se realizó una segmentación de K-Means y se determinó de acuerdo a las distancias y la silueta que la mejor división está en dos clúster.

Palabras claves: Componentes principales, K-Means, Segmentación

## Tabla de Contenidos

7

<a href="#"><u>Capítulo 1 Introducción e información general</u></a>	1
<a href="#"><u>Capítulo 2 Planteamiento del Problema</u></a>	2
<a href="#"><u>Objetivos</u></a>	3
<a href="#"><u>Objetivo General.</u></a>	3
<a href="#"><u>Objetivos Especificos.</u></a>	3
<a href="#"><u>Justificación</u></a>	4
<a href="#"><u>Antecedentes</u></a>	5
<a href="#"><u>Capítulo 3 Marco Teorico / Conceptual.</u></a>	8
<a href="#"><u>Entorno Financiero y Productivo</u></a>	8
<a href="#"><u>Nuestro Enfoque</u></a>	11
<a href="#"><u>Entorno Socioeconomico</u></a>	12
<a href="#"><u>Creacion de Perfil Nicho de Mercado</u></a>	14
<a href="#"><u>Capítulo 4 Marco Metodologico.</u></a>	18
<a href="#"><u>Análisis Descriptivo limpieza de bases</u></a>	19
<a href="#"><u>Análisis de componentes principales</u></a>	23
<a href="#"><u>Metodologia K Means</u></a>	23
<a href="#"><u>Capítulo 5 Analisis y Resultados.</u></a>	27
<a href="#"><u>Conclusiones</u></a>	33
<a href="#"><u>Recomendaciones</u></a>	34
<a href="#"><u>Bibliografía</u></a>	35
<a href="#"><u>Apéndice</u></a>	37

## Lista de tablas

8

Tabla No 1 Principales Magnitudes Financieras de la IFD Agricolas 2000.	10
Tabla No 2 Variables sociodemograficas y financieras .	19
<a href="#"><u>Tabla No 3 variables se excluyen del estudio, por inconsistencias.</u></a>	21
<a href="#"><u>Tabla No 4 Variables se excluyen del estudio, por altaconcentración.</u></a>	21
<a href="#"><u>Tabla No 5 Numero de datos atipicos.</u></a>	22
<a href="#"><u>Tabla No 6 % Acumulación de varianza de los componentes.</u></a>	28
<a href="#"><u>Tabla No 7 Correlación de las Variables.</u></a>	28
<a href="#"><u>Tabla No 8 Indicadores promedio por Cluster Perfiles.</u></a>	32



## Lista de gráficos

9

<a href="#"><u>Grafico 1. Centroides Iniciales.</u></a>	25
<a href="#"><u>Grafico 2. Reducción de variables.</u></a>	25
<a href="#"><u>Grafico 3. División tamaño cluster.</u></a>	27
<a href="#"><u>Grafico 4. Silueta SPSS MODELER.</u></a>	30
<a href="#"><u>Grafico 5. Resultado Modelo SPSS MODELER.</u></a>	31

## **Capítulo 1**

### **Introducción e información general**

El trabajo tuvo como objetivo presentar el uso de la técnica de análisis de componentes principales con los indicadores financieros de clientes de caña panelera de una entidad financiera y posteriormente segmentación de K-Means y de acuerdo a la similitud de indicadores intragrupos pero muy diferentes entre grupos interpretar los clúster.

El desarrollo de este trabajo fue muy importante para la entidad bancaria ya que la herramienta brinda un perfil para otorgamiento de crédito a clientes que de acuerdo al análisis son menos riesgosos y son sujetos a un crédito de una manera más rápida.

El documento consta de Capítulo I, organización del documento. Capítulo II motivo por el cual se realiza la investigación, propósito de lo que se quiso lograr. Capítulo III, es el marco teórico donde se realizó la revisión de documentación como investigaciones, papers, revistas, tesis, documentales entre otros. En el capítulo IV, se definirá la población objetivo, selección de las variables. En el capítulo V, se encuentran los resultados luego de aplicar el análisis de componentes principales y la técnica de segmentación de K-Means. Por otro lado se muestran las conclusiones y recomendaciones y por último se presenta las referencias bibliográficas y anexos.

## **Capítulo 2**

### **Planteamiento del Problema**

La conciencia del ahorro aun esta poco arraigado en nuestra sociedad es un dilema diario cuando estamos inmersos en una sociedad consumista y requerimos de bienes o servicios que pueden ser muy necesarios como también muy innecesarios y superfluos, pero la simple satisfacción de obtener lo que se desea siempre hace pensar en diferentes modos de obtener lo que se requiere.

Aquí es donde nacen los bancos, su función principal es brindar soluciones financieras al alcance de todo el que quiera y pueda pagar una obligación y como retribución pagamos el favor con dinero o llamado intereses.

Hasta este punto no hay nada fuera de lo normal, si tienes como pagar el dinero que te prestan y adicional la retribución por tu préstamo, eres el feliz ganador de un crédito en cualquiera de sus formas, ya sea tarjeta de crédito, crédito de consumo, crédito hipotecario, crédito de libranza, etc., existen productos para todos los gustos. Pero la pregunta es la siguiente: ¿Cómo saben los bancos o entidades bancarias para conocer o saber a quién se le puede prestar dinero y a quien no? Si no estás inmerso dentro del sector financiero es como lanzar una piedra al aire tienes el 50% que te caiga en la cabeza, todo depende de los factores que afecten como el viento y la dirección y tendrás más probabilidad de que si te caiga en la cabeza.

Con las entidades financieras es igual, al inicio tienes un 50% de que te presten o no, pero la balanza la puedes inclinar a tu favor de acuerdo a tu comportamiento en el sector, si tienes mal comportamiento no eres viable de crédito y si tienes buen comportamiento eres viable para prestarte su dinero.

Por otro lado los bancos en los últimos años le dan una gran importancia al conocimiento y fidelización de los clientes, lo relevante es saber lo que el cliente quiere y no lo que el banco necesita. Anteriormente las entidades financieras realizaban perfilamiento de clientes sin técnicas adecuadas que les ayudara a segmentar con diferentes variables, solo se usaba análisis descriptivos.

Lo que se busca con este proyecto es perfilar los clientes del destino caña panelera y quienes son aptos para otorgamiento de crédito de acuerdo a unas variables financieras y entrarían por un análisis más rápido para desembolso de crédito y quienes no sean aptos continuarían el proceso regular con los tiempos estipulados por el banco.

## **Objetivos**

### **Objetivo General**

Generar una metodología estadística que proporcione un perfil adecuado para la aprobación de créditos financieros destinado a productores de caña panelera.

### **Objetivos Específicos**

- Determinar las principales variables financieras relacionadas con los perfiles de los clientes de caña panelera.
- Desarrollar una metodología de análisis para el perfil de usuarios de caña panelera.
- Validar y analizar el modelo de análisis de los perfiles de clientes de caña panelera.
- Pronosticar dado el perfil de un productor de caña panelera, si le será otorgado el crédito solicitado

### **Justificación**

En la actualidad económica el principal objetivo es no naufragar en la economía de escala en la que estamos inmersos y que son necesarios los riesgos y los retos que derivan del afán de sobresalir en el consumismo sea cual sea su actividad económica.

Si bien antes los créditos financieros no eran tan de fácil acceso como hoy, los favorecidos eran torturados con unas tasas muy altas, porque las entidades debían cubrir sus posibles pérdidas de aquellos clientes que no cumplían con sus obligaciones con los pagos oportunos y completos de los clientes fieles y buenos, lo anterior solo tiene una explicación, no existía metodologías de estudios de créditos ni perfilamiento de clientes.

Es por esto que es completamente necesario para una entidad financiera realizar un análisis de crédito y perfilamiento de clientes de acuerdo a sus comportamientos en el sector económico más aun cuando en la actualidad te toca competir con los demás bancos y no es posible por tasa porque ya este tema se encuentra regulado por el banco central,

solo queda arriesgar a un otorgamiento con el mínimo riesgo y tratar de conocer mediante cifras e indicadores a nuestros posibles clientes, mediante técnicas de segmentación que logren un resultado favorable que nos ayude a obtener un perfil de cliente menos riesgoso o que no tengan experiencia crediticia poder determinar por medio de variables sociodemográficas o financieras un modelo de perfil adecuado para crédito.

### **Antecedentes**

Se habla que en la edad media puntualmente en Grecia y Roma existía el crédito privado donde el acreedor entregaba al deudor un bien para su uso y si el deudor no cumplía con su obligación el bien pasaba de nuevo a manos del acreedor.

La palabra crédito viene del latín crédito de credere o tener confianza y es el principio del pasivo. Aunque la base del crédito es la confianza para los casos actuales es la esperanza de recuperar lo prestado porque siempre existirá el temor o la desconfianza de no lograr recuperar lo prestado. En el momento cero se parte de la calidad moral o buena intención del deudor en cumplir con su obligación y esta es la mayor garantía pero también se pueden solicitar garantías de menor valor como bienes muebles o inmuebles.

Ya sea cuando un negocio logra aumentar sus operaciones o inicia desde la nada y no cuenta con el dinero suficiente recurre a un crédito y el uso de este será la causa del éxito o del fracaso de la empresa, y es por esta última razón que se hace imprescindible que las

empresas que otorgan créditos realicen un modelo intuitivo de aprobación de préstamos, el objetivo final es mitigar la posible pérdida de dinero por incumplimiento del deudor.

En los créditos existen muchas variables que definen si es viable o no una aprobación, una de ellas es la tasa de interés, como lo mencione anteriormente al no existir un adecuado análisis de los clientes los costos por intereses eran muy elevados lo que disminuía el porcentaje de la población apta para otorgamiento de crédito y la otra variable es el apetito del riesgo y es cuanto estoy dispuesto a perder como institución financiera si mis clientes no me pagan, la relación entre estas dos variables están estrechamente relacionadas porque dependiendo de mi rango de perdida y mi cantidad proyectada de préstamos puedo proyectar a que tasa puedo prestar para que mi porcentaje de ganancia sea positivo.

En la actualidad es tanta la demanda de soluciones financieras que si no estuviera regulado los intereses estarían por las nubes, y por eso las entidades financieras permanecen en constante lucha para ganar la mayor cantidad de clientes como les sea posible inclusive quitándoselos unos a los otros, ofreciendo tantas modalidades y beneficios a los posibles nuevos clientes que su competencia radica en el buen servicio y para poder ofrecer esto se necesita asumir ciertos costos que los debe garantizar la dinámica del negocio y por esta razón sus niveles de perdida deben estar mínimos porque las tasas de interés deben ser muy competitivas y sus modelos de perfilamiento deben estar a la vanguardia de la evolución de la economía entrando inclusive a mercados pocos

explorados como los agropecuarios de los cuales deben sacar la mayor cantidad de información, pues si generar un perfil a una persona que se encuentra en el ámbito laboral normal es más aun difícil lograr perfilar a un productor de caña panelera que no tiene historia crediticia y que su único soporte es la historia de su producto en términos comerciales, de cuanto vale sus tierras, nivel de producción, valor de venta e indicadores financieros.



## **Capítulo 3**

### **Marco Teórico / conceptual**

#### Entorno Financiero Y Productivo

En los años ochenta y noventa de acuerdo a las reformas agrarias de los gobiernos latinoamericanos se estableció un objetivo claro de favorecer al sector agrícola, política que fue implementada y puesta en funcionamiento, salvo que para inicios de la década de los años noventa fue materia de fuertes cuestionamientos hacia los bancos y promotores del plan de desarrollo rural y es que el objetivo de favorecer a los pequeños productores rurales no se había cumplido ya que los esfuerzos fueron enfocados en la favorabilidad de los sectores más modernos de la agricultura comercial. (Romel Acevedo y Gabriel Delgado, 2001).

De alguna manera esta forma de favorabilidad se debía a la poca profundización en la producción agrícola del pequeño productor, no es fácil recolectar información sobre sus estados financieros, si son llevados con la misma delicadeza y organización que una gran empresa, el riesgo era fuerte para no tener en cuenta este nicho de población.

No todo fue cerrar puertas y oportunidades, hay un ejemplo a resaltar es el BANRURAL de Guatemala, pionero en la adjudicación de préstamos y servicios de créditos a los sectores agropecuario, artesanal, comercial, industrial y demás sectores económicos del país, lo interesante es que este banco es que no se limitó a la política económica de la época, encamino sus esfuerzos hacia otros destinos financieros como banca de segundo piso, realizando operaciones de financiamiento, refinanciamiento y cofinanciamiento a intermediarios lo que trae como consecuencia que estos últimos sirvan como puente en la adjudicación de recursos financieros a las micro, pequeñas y medianas empresas del agro y la trajo al banco un crecimiento exponencial en su nivel de activos y utilidades.

El sector se dinamizo y se crearon bancos especializados en el agro y bancos de segundo nivel que coadyuvaban en la consecución del objetivo de los gobiernos de la zona.

En el siguiente cuadro se muestran por países los bancos especializados en el agro y los países que cuentan con bancos de segundo nivel enfocados hacia el agro.

Tabla No 1 Principales Magnitudes Financieras de la IFD Agrícolas 2000

Principales magnitudes financieras de las IFD agrícolas en 2000					
Institución (País)	Activos US\$ M	Préstamos US\$ M	Patrimonio US\$ M	Cartera agrícola %	Contribución al financiamiento agrícola %
<b>Bancos agrícolas especializados</b>					
INDAP (Chile)	185	153	112	100	
Banco Agrario (Colombia)	1.508	338	85	41,1	
Banco Nacional de Fomento (Ecuador)	119	37	12	91	
Banco de Fomento Agropecuario (El Salvador)	190	89	17	100	
Banco de Desarrollo Rural (Guatemala)	345	141	24	42	36,40
Banco Nacional de Desarrollo Agrícola (Honduras)	53	13	2	78	3,60
Banco Nacional de Crédito Rural (México)	3.626	2.826	182	100	2,60
Banco de Desarrollo Agropecuario (Panamá)	192	104	81	100	21,20
Banco Nacional de Fomento (Paraguay)	325	199	38	62	41,50
Banco Agrícola (Rep. Dominicana)	241	146	27	100	39,50
<b>Banca de segundo piso especializada y multisectorial</b>					
Nacional Financiera (Bolivia)	160	122	7	50,00	16,00
Banco Nacional de Desenvolvimento Económico y Social (BNDES) (Brasil)	51.509	38.333	6.027	8,30	7,20
FINAGRO (Colombia)	1.047	789	134	100,00	
Corporación Financiera Nacional (Ecuador)	1.072	867	196	27,90	
Banco Multisectorial de Inversiones (El Salvador)	575	386	139	11,00	
FIRA-Banco (México)	10.718	2.126	2.621	59,90	32,20
Financiera Nicaragüense de Inversiones (Nicaragua)	86	78	54	53,70	16,70
Corporación Financiera de Desarrollo (Perú) (COFIDE)	1.325	745	393	5,20	
Banco de Desarrollo Económico (Puerto Rico)	950	114	95	17,40	
FONDAFA (Venezuela)	265	99	219	100,00	7,70
<b>Banca pública multisectorial, comercial y de desarrollo</b>					
Banco de la Nación (Argentina)	17.873	10.388	1.438	36,40	61,30
Banco de la Prov. de Bs. As (Argentina)	14.664	10.876	1.461	7,57	
Banco do Brasil	70.486	34.563	4.058	70,00	60,00
Banco do Nordeste (Brasil)	4.486	3.331	508	60,00	32,60
Banco del Estado (Chile)	9.559	5.994	598	4,80	14,30
Banco Nacional (Costa Rica)	2.694	1.312	107	16,00	56,00
Banco de Crédito Agrícola de Cartago (Costa Rica)	468	128	28	8,50	3,90
Banco de la República Oriental (Uruguay)	4.757	3.593	499	30,00	72,00

Fuente: ALIDE - Base de Datos.

Para nuestro caso específico la política de crédito agrícola se enmarca dentro de los lineamientos de la ley 16 de 1990 se estableció el Sistema Nacional de crédito agropecuario y se creó el fondo para el financiamiento del sector agropecuario FINAGRO como entidad especializada de segundo piso que canaliza sus recursos a través del sector bancario.

También tenemos banco de primer piso el BANCO AGRARIO, constituido en el año 1998 al ser liquidado la caja Agraria, sus operaciones ascienden a US338 millones de dólares y su directriz es la financiación de las actividades rurales, agrícolas, pecuarias, pesqueras, forestales y agroindustriales.

### **Nuestro Enfoque**

Ya conocemos la evolución del sistema financiero es hora de conocer sobre el producto y sus derivados, la caña panelera es la materia prima esencial en la elaboración de la panela, de acuerdo a estudio realizado en el año 2014 Estrategias para aumentar las utilidades de la producción panelera en Colombia por Carolina Melo, describe la composición del sector encontrándose a los pequeños productores en el 92% de la producción y tan solo el 8% restante son las medianas y grandes empresas, esto solo dice

que en Colombia tenemos un gran potencial en convertirnos en el país número 1 de exportación de panela en el mundo, en la actualidad ocupamos el segundo lugar después de la India.

Pero que nos detiene? El obstáculo más grande es la tecnificación e industrialización de la producción de la panela y de la caña panelera, tecnologías limpias que contribuyan con el medio ambiente y genere utilidades rentables para sus productores, Grados Brix (Tecnologías Limpias, 1992).

Para lograr el objetivo de la tecnificación de los micros y pequeños productores se hace necesario que el gobierno nacional destine partidas presupuestales para este nicho y que el banco especialista en el agro BANCO AGRARIO, abra sus brazos y encuentre una forma de fomentar los préstamos a una tasa exequible que contribuya y no deteriore los aspectos económicos de estos microempresarios, razón por la cual mi intención es dar una posible solución mediante un modelo estadístico de otorgamiento de crédito y de alguna forma contribuir al crecimiento del agro en el producto de caña panelera.

### **Entorno Socioeconómico**

La dinámica de la economía actual está marcada y centrada en el sector financiero, conseguir una aprobación para obtener un crédito cada día es más fácil, pues las instituciones financieras están más asequibles a personas que no tengan un gran historial

crediticio porque uno de los objetivos es la bancarización como en su anuncio lo realiza Andrés Murcia Pabón 2007 DETERMINANTES DEL ACCESO A CREDITO DE LOS HOGARES COLOMBIANOS “Durante los últimos años diferentes sectores del gobierno y la academia han discutido sobre la necesidad de ampliar el acceso a los servicios financieros para la mayoría de hogares posibles y así lograr mejores condiciones en términos de oportunidades y bienestar de la población....”, ahora ya teniendo el objetivo claro no se debe discriminar el nivel económico ni productivo de ninguna población y es donde se abre una oportunidad para el sector agrario.

Inicia con aquellos agricultores con productos atractivos y comerciales como el café y la ganadería, con organizaciones bien estructuradas que respaldan a estos productores segregados en zonas específicas del territorio nacional y que con ayuda de las agremiaciones logran conseguir o acceder a préstamos financieros logrando tener un capital de trabajo acorde a sus necesidades productivas.

De acuerdo al resultado del estudio de Miranda y Paredes (2006) los principales factores de negación de los créditos son los bajos ingresos y falta de historia crediticia y de aquí nace la tesis que es necesario realizar salir del círculo y explorar nuevas alternativas que permitan el acceso a servicios financieros a aquellos productores de caña panelera donde su insumo se convierte en un producto de primera necesidad.

Para generar un nuevo portafolio se debe conocer muy bien el producto de la caña el cual está muy bien explicado en el documento de DESARROLLO DE LAS PEQUEÑAS

INDUSTRIAS DE LA CAÑA DE AZUCAR EN IBEROAMERICA realizado por Fabio Cesar Dasilva, Marco Antonio Azeredo Cesar y Carlos Gregorio Hernández Díaz-Ambrona (2010). A lo largo del tiempo la industria tuvo sus altos y bajos en cuanto a la relación económica productiva y para controlar esos desajustes se inició con el proceso de disciplinar la producción y controlar el mercado interno y externo, por esta iniciativa fue el inicio del desarrollo administrativo al interior de los productores de la caña, inicios que llevaron a tener un control más adecuado de los ingresos y costos productivos lo que para este proyecto es vital tener acceso a esta información pues con ella se puede generar el perfil adecuado para la exploración financiera a este nuevo nicho de mercado.

### **Creación Perfil Nicho Mercado**

El empleo de análisis multivariado ayuda a mejorar la interpretación de problemas con muchas variables, esta metodología no es nueva en Colombia, hay muchos trabajos que se han realizado en diferentes ámbitos utilizando la reducción de datos y la dimensionalidad de objetos de investigación. Johnson y Wichern (1998), el análisis de componente principal suele ser muy útil para identificar variables que estén altamente correlacionadas o que tengan grados de asociación significantes, lo que llamamos multicolinealidad entre variables, en caso se detectaría podría omitir dicha variables esto para evitar la redundancia de información en el estudio.

Ospina y Lema (2005) emplea análisis de componentes principales (ACP) para encontrar las relaciones entre variables y clasificar los impactos por grado de influencia. Frente a la reducción de dimensiones de un gran número de variables, Lindsay I. Smith (2002), nos indica que las componentes principales también que grupos explican mayor o poca variabilidad del conjunto de variables en estudio, por ello toma interés los indicadores de similaridad entre las variables.

Hart, R (1998) expresa que los métodos estadísticos multivariados permiten operacionalizar el sentido heurístico de la teoría de sistemas, aplicada al problema de clasificar y tipificar. Laurebt (1998) afirma que para llegar a un grado de simplicidad en la validación es necesario tener en cuenta el entorno socioeconómico de la región, dado que este influye y es necesario hacer uso de las características encontradas al analizar las jerarquías de los sistemas agrícolas. Por otro lado Stefos et al (2011) en su estudio manifiesta que el análisis estadístico multidimensional se utiliza con el fin de mostrar las principales y más importantes criterios de diferenciación y la clasificación en grupos de las personas investigadas. Los métodos que utiliza son el análisis factorial de correspondencias múltiples que muestra los criterios de diferenciación y la agrupación jerárquica que define los grupos de las personas debido a sus características comunes. En este mismo sentido Dixon et (2002), Pardo et (2002), Visauta et (2003) definen los métodos de análisis multivariado como un conjunto de técnicas de análisis de datos que permiten al investigador testar la utilidad conjunta de datos que se emplean en la explicación de los fenómenos de interés analítico como económico, social, ambiental, desarrollo sostenible entre otros. El ACP Johnson (2000) es de utilidad porque ayuda a



revelar anomalías en los datos y descubrir datos atípicos, ayuda al investigador a detectar subgrupos de acuerdo al comportamiento común de la varianza de las variables respuesta y por último ayuda a disminuir las variables en un número menor de componentes principales que expliquen el mayor porcentaje de varianza de los datos originales.

Hoy en día para el análisis de datos existen diferentes técnicas que nos pueden ayudar a entender el comportamiento de los mismos, dentro de ellas se puede destacar una de las técnicas más utilizadas la cual es conocida como la Minería de Datos, esta técnica consiste en la iteración de una serie de pasos los cuales son: Selección, Limpieza, Transformación, Evaluación y Representación del Modelo. Berndt y Clifford (1996), estima que el suministro de datos del mundo se duplica cada 20 meses, lo anterior implica un crecimiento excesivo en el volumen de datos que se maneja tanto en la comunidad científica como en los sectores productivos de la economía que ha sobrepasado la capacidad humana de analizar, resumir y extraer conocimientos a tales cantidades de datos, por lo anterior se hace necesario una nueva generación de herramientas capaces de automatizar el análisis de los datos almacenados. El conjunto de estas herramientas lo estudia un nuevo campo de investigación llamado minería de datos Fayyad et al (1996). La minería de datos se ha convertido en una herramienta indispensable para la toma de decisiones financieras, producción, organización y demás factores en las empresas que de cierta manera las hacen más competitivas. Inmon, (1992), afirma que la minería de datos es también un conjunto de datos integrados orientados a una materia, que varían con el

tiempo, y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración.

Por otro lado Simón, 1997 concluye que la minería de datos es un conjunto de técnica de análisis de datos que permiten extraer patrones, tendencias y regularidades para describir y comprender mejor los datos y extraer patrones y tendencias para predecir comportamientos futuros. Bonilla y Bermúdez (2013) Utilizando el test de significancia Chi-cuadrado, con datas extensas, analizan las diferencias por género, el test les afirma que existe una notoria brecha de ingresos y desigualdad de oportunidades. Aguirre (2011) desarrolla modelos ordinales logísticos, con variables dependientes no estandarizadas e independientes estandarizadas, asimismo, identifica cuatro variables de un volumen grande de las mismas. Existen variedades de minería de datos los clúster se han vuelto tendencia en el mundo y es un algoritmo desarrollado por IBM e implementado en el IM, componente del DWE, que resuelve automáticamente los problemas de definición de métricas de distancia / similitud, proporcionando criterios para definir una segmentación óptima (Grabmeier, & Rudolph, 1998), en este estudio se optara por clusterizacion de variables para una mejor discriminación de los datos.

## **Capítulo 4**

### **Marco Metodológico**

Para los análisis se utilizó información de fuentes de datos de clientes de caña panelera años 2008-2019 de una entidad bancaria, la investigación es descriptiva ya que se cuenta con datos reales. Nuestros clientes objetivos son aquellos que tienen marca de pago sí, es decir que ya han realizado algún abono a capital intereses u otros conceptos, aquellos clientes que están en periodo de gracia no se tuvieron en cuenta para el análisis, el periodo de tiempo tomado es a partir del año 2008.

El enfoque del trabajo es cuantitativo ya que se busca determinar un perfil a partir de unas variables financieras y posteriormente diferenciar en grupos los mejores clientes.

El método de reducción de variables consistió en el análisis de componentes principales para la explicación de variabilidad de las variables (ACP), luego se realiza con la

metodología de K-Means grupos diferenciadores pero con características en común en sus indicadores financieros.

- 4.1. Análisis descriptivo limpieza de base.
- 4.2. Análisis de Componentes Principales.
- 4.3. Metodología K-Means.

#### **Análisis descriptivo limpieza de base**

- Base: Contiene 5.407 obligaciones y 46 Variables sociodemográficas y financieras.

Tabla No2 (Fuente Propia)

VARIABLES	DESCRIPCIÓN
No	Numero consecutivo de Obligación
IND_Apalancamiento	Indicador Porcentual: $(\text{Pasivo Total} / \text{Patrimonio Total})$
IND_Endeudamiento	Indicador Porcentual: $(\text{Pasivo Total} / \text{Activo Total})$
Cobertura_Interes	Indicador Porcentual: $(\text{Utilidad de Operación} / \text{Intereses})$
Razon_Corriente	Indicador Porcentual: $(\text{Activo Corriente} / \text{Pasivo Corriente})$
Rentabilidad_Sobre_Activos	Indicador Porcentual: $(\text{Utilidad Neta} / \text{Activo Total})$
Rentabilidad_Sobre_Patrimonio	Indicador Porcentual: $(\text{Utilidad Neta} / \text{Patrimonio Total})$

	/Patrimonio Total)
IND_Desarrollo_Nuevos_Productos	Indicador Porcentual: (Pasivo corriente/ pasivo Total)
ICV	Indicador de Mora
pdi_final	Indicador de Perdida dado el Incumplimiento de acuerdo a la garantía.
experiencia_actividad	Si tiene experiencia en la Actividad (S)
tiempo_actividad	Tiempo en la actividad en años
experiencia_BA	Si tiene Experiencia en el Banco (S)
calidad_experiencia	Marcación interna
costo_ventas	Valor de los costos de Ventas anuales
utilidad_neta	Valor Utilidad Anual
ingresos	Valor de Ingresos Anual
edad	Edad del cliente
fuentes_alternativas	si tiene otros ingresos adicionales
puntaje_CIFIN	Puntaje Score historia cliente
tvrc_modelo	Tipo de Modelo
puntaje_VRC	Puntaje Score Interno Banco
Marca_Victimas	Si es Victima y tomo Credito o tenia un credito y se declaro victima.
fo_clase	Tipo de Cartera según Superfinanciera 1: Comercial 2: Consumo 3: Vivienda 4: Microcrédito.
destino_agrupado	Destino del Crédito
COSECHA	Fecha de Desembolso
fo_calificacion	Calificación del Cliente
fo_monto_des	Monto Desembolsado
fo_saldo_cap	Saldo de capital activo
fo_dias_mora	Dias de Mora del Crédito
fo_reestructurada	Si es Crédito Reestructurado
MUNICIPIO_INVER	Municipio de Inversión
regional	Regional donde se solicita Crédito.

zonal	Zonal donde se solicita el Crédito
DESTINO_FINAL	Destino Final Agropecuario o No agropecuario
tipo_de_productor_agrupado	Tipo de productor: pequeño, Mediano, Grande.
saldo_m30	monto de cartera Vencida Activa
marca_pago	Marca de Pago SI o NO
CLASIF_ESTATUTOS_DNP	Si es Rural o Urbano
tipo_asesor	Asesor
Banca Cliente	Banca Agropecuaria, Empresarial, Microfinanzas u Oficial.
PE_Clase_cartera	Monto de Perdida Esperada
Record_Count	casos
%_Perdida_Esperada	Indicador de Perdida
B_M	Marcación Buenos o malos de acuerdo a la Perdidad esperada.

- Limpieza de datos

Las siguientes variables se excluyen del estudio, dado que se evidencia inconsistencias en sus registros:

Tabla No3 (Fuente Propia)

costo_ventas
utilidad_neta
ingresos

Las siguientes variables cualitativas se excluyen del estudio dado que se evidencia una alta concentración en uno o pocas categorías de la variable, por consiguiente, no permite una eficiente agrupación de los registros:

Tabla No4 (Fuente Propia)

ICV
experiencia_BA
experiencia_actividad
Marca_Victimas
fo_clase
fo_reestructurada
fo_calificacion
regional
DESTINO_FINAL
CLASIF_ESTATUTOS_DNP
tipo_asesor
Banca Cliente

Con las variables cuantitativas, se realizan tablas de frecuencia, boxplot y se eliminan registros que no alteren la población, pero si se consideran datos atípicos por tener una diferencia considerable con el 3 cuartil (75%) de la variable. El anterior procedimiento se realiza, con el objetivo de disminuir variabilidad para estandarizar las variables (Razón Corriente, Indicador de crecimiento de ventas y Cobertura de Intereses). Se eliminaron 475 registros de 5407, es decir el 8.7% de la base de estudio. Los criterios de eliminación son los siguientes:

Tabla No5 (Fuente Propia)

Variables	Datos Atípicos-Tabla de Frecuencia		Estandarización
	Criterio	#Eliminados	
IND_crecimiento_ventas	>500	15	99
IND_Apalancamiento	>2	9	
IND_Endeudamiento	>0.65	4	
Cobertura_Interes	>150	57	126
Razon_Corriente			131
Rentabilidad_Sobre_Activos	>0.8	8	
Rentabilidad_Sobre_Patrimonio	>0.7	11	
IND_Desarrollo_Nuevos_Productos	>1000	4	
puntaje_CIFIN			
puntaje_VRC	>280000000	5	
fo_monto_des	>240000000	3	
Ffo_saldo_cap	>1000	3	
tiempo_actividad			

### Análisis Componentes Principales

El objetivo del análisis de componentes principales (ACP) es resumir un número alto de variables continuas en un nuevo grupo sin perder cierta parte de información original significativa para ello usamos el software RStudio y utilizamos los indicadores financieros los enfoques que se validaron son:

1. Análisis de matriz de correlación: La finalidad de la matriz de correlaciones es comprobar si sus características son las adecuadas para realizar el análisis factorial.
2. Varianza: Explicada a través de los eigenvalores (es decir la proporción total de varianza que explica ese factor para las variables consideradas como grupo).



## Metodología K-Means

Obteniendo con el ACP los indicadores que mejor me explican la varianza de los datos se procede a realizar por medio del software SPSS MODELER análisis con metodología de agrupamiento que consiste en una técnica de acuerdo a la distancia y su grado de similitud de sus elementos dentro de cada clúster este método de agrupamiento es no jerárquico, ya que debe realizarse una asignación inicial de un número de grupos de referencia.

Los pasos son:

1. Se ingresan los indicadores financieros y mediante distancia euclidiana determina similitud en las observaciones (los datos están representados por los promedios de las variables del individuo asignados a un grupo y estos serán los centroides y cada individuo se clasificara en el centroide más cercano de acuerdo a la distancia.

Formula:

$$D_{ic} = \sqrt{\sum (X_{ij} - X_{ic})^2}$$

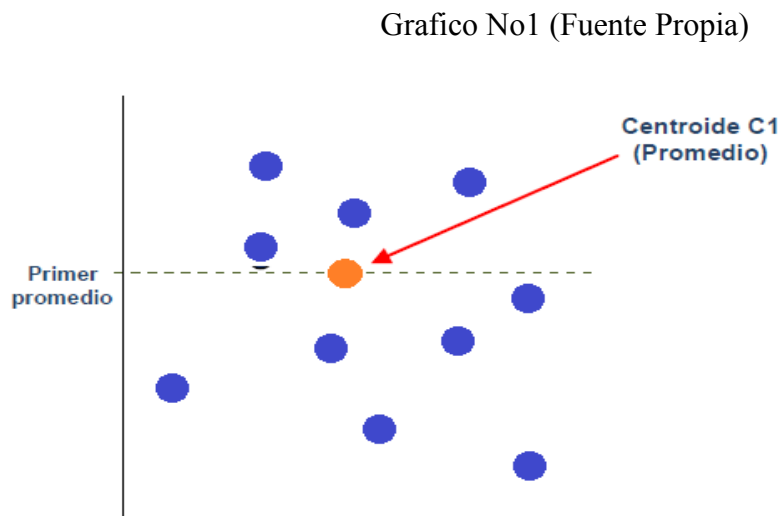
Dónde: i: elemento i-ésimo

j: variable j-ésimo

c: centroide del clúster

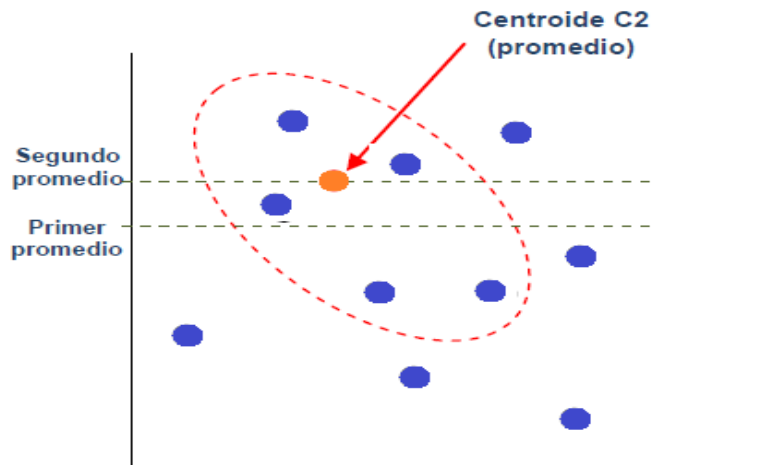
2. Luego realiza nuevamente iteraciones hasta que las observaciones sean muy similares entre ellas pero diferentes entre grupos.

Primer Centroide: nos muestra el centroide inicial que se tomó como referencia, que consiste en el promedio de las variables de toda la población.



Segundo Centroide: se observa que una vez que se creó el nuevo grupo con respecto a las distancias más cercanas de los individuos hacia el centroide referencial, el nuevo centro se recalcula para el grupo, y así sucesivamente el método K-Means se recalculará hasta crear tantos grupos sean que cumplan los criterios de similaridad.

Grafico No2 (Fuente Propia)



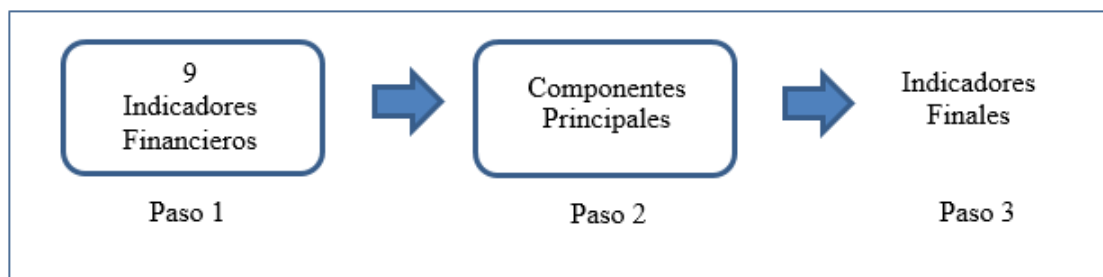
3. Esto se le debe realizar asignando el número de clúster hasta que se obtenga la mayor discriminación entre grupos y cuando la medida de la silueta de cohesión y separación este por encima del 50%.

## Capítulo 5

### Análisis y Resultados

#### Análisis de Componentes Principales (reducción de variables)

Grafico No3 (Fuente Propia)



En grafico nos muestra el proceso de cómo reducir las variables en la dimensión de indicadores financieros con el objetivo de facilitar la interpretación de todas las variables que explicarían el comportamiento de los clientes.

El Paso 1 nos indica que partimos de 9 variables financieras, en el Paso 2 aplicamos la técnica de análisis de componentes principales (ACP) para la reducción de variables el cual se desarrolló de la siguiente manera:

Analizando las variables en el software estadístico RSTUDIO con la cantidad de factores que se obtuvo con auto valores mayores a 1, se obtuvo que la explicación de los factores hacia las variables es muy poco, por tal motivo se realiza la reducción de variables seleccionando las que aporten a la variabilidad del conjunto de datos y a su vez reduzca dimensionalidad, para esto se debe seleccionar las dimensiones con valores propios mayores a 1.

Se selecciona 4 componentes que acumulan el 67.63% de varianza:

Tabla No6 (Fuente Propia)

Componentes	Valores Propios	% de varianza	% acumulado de Varianza
1	2,526	28,064	28,06
2	1,453	16,143	44,21
3	1,111	12,342	56,55
4	1,002	11,076	67,63

Paso 3: Para seleccionar las variables que aportar más a la variabilidad, se toma el criterio de correlación de las variables y el componente.

Tabla No7 (Fuente Propia)

	Dim. 1	Dim. 2	Dim. 3	Dim. 4
IND_crecimiento_ventas	0,048	0,406	0,195	0,069
IND_Apalancamiento	0,922	-0,163	0,075	-0,044
IND_Endeudamiento	0,938	-0,193	0,089	-0,027
Cobertura_Interes	-0,378	0,467	0,132	0,029
Razon_Corriente	-0,371	0,296	0,489	-0,43
Rentabilidad_Sobre_Activos	0,349	0,726	-0,165	0,013
Rentabilidad_Sobre_Patrimonio	0,593	0,609	0,043	0,1
IND_Desarrollo_Nuevos_Productos	-0,133	0,144	-0,062	0,171
puntaje_VRC	-0,145	-0,025	0,344	0,874

#### Grupo final

1. Indicador de Apalancamiento
2. Indicador de Endeudamiento
3. Razón Corriente
4. Rentabilidad sobre Activos
5. Puntaje VRC

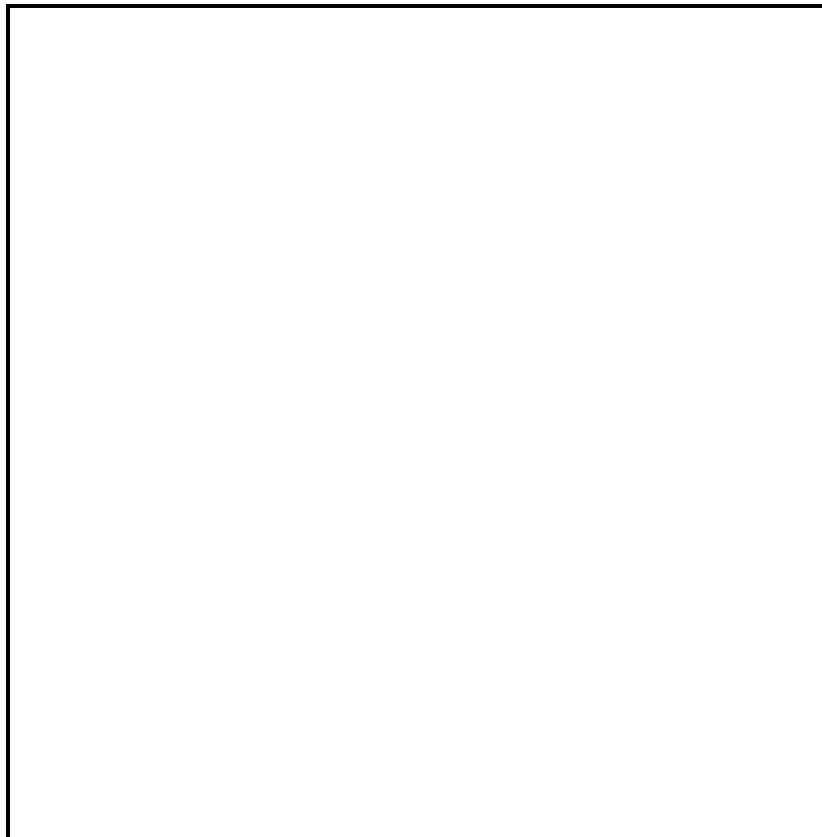
#### Clúster (Método K-Means)

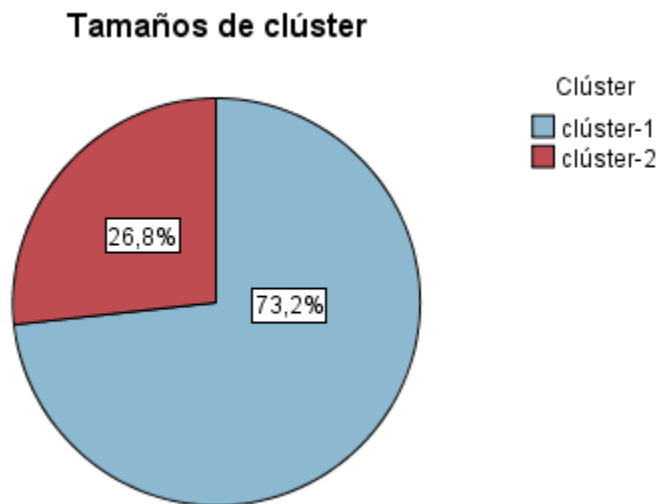
Con 4 variables finales (Indicador de Apalancamiento, Indicador de Endeudamiento, Razón Corriente y Puntaje VRC), se aplica en el software SPSS Modeler la técnica de K-Means, la cual realiza una serie de iteraciones.

El gráfico muestra el total de la población en estudio 4.932, el 26% dista de el 73% es decir que son muy diferentes entre grupos de acuerdo a los indicadores financieros pero intra grupos tienen similitudes de acuerdo a la distancia euclídea (Es la raíz cuadrada de la suma de los cuadrados de las diferencias entre los valores de los elementos y esta es la medida por defecto que suele usarse para datos de intervalo en el k means).

La división más apropiada para el grupo es la división es de 2 clúster, lo que nos indica que los segmentos son más diferenciados entre ellos.

Grafico No4 (Fuente SPSS MODELER)





<b>Tamaño del clúster más pequeño</b>	1323 (26,8%)
<b>Tamaño del clúster más grande</b>	3609 (73,2%)
<b>Cociente de tamaños: De clúster más grande a clúster más pequeño</b>	2,73

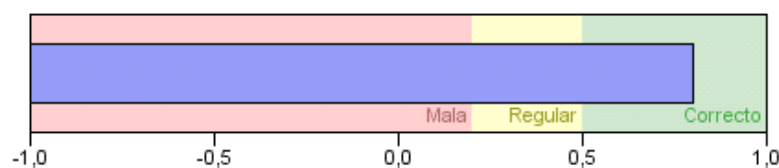
El k-Means nos muestra la silueta y es la forma o la pureza que tienen los segmentos entre ellas y está compuesta por la distancia interior o inercia inter clase y la inercia intra clase, la cohesión y separación de los datos es adecuada.



### Resumen del modelo

<b>Algoritmo</b>	K-medias
<b>Entradas</b>	4
<b>Clústeres</b>	2

### Calidad de clúster



Observamos la comparación de variables de acuerdo a su promedio en los dos clúster y podemos determinar una segmentación de acuerdo a los resultados donde el primero es el más adecuado porque se caracteriza:

1. Clientes que tienen un apalancamiento en promedio de 14%, es decir que no se han financiado altamente con deuda (entre menor sea el indicador es mejor).
2. Clientes con endeudamiento en promedio del 12%, este indicador mide el grado de endeudamiento con relación a sus activos y lo que indica es que los clientes están financiando 12% con recursos ajenos y 88% con recursos propios. (Entre menor sea el indicador es mejor).

3. Clientes con indicador corriente promedio 7.59 que es la capacidad que se tiene para cumplir con las obligaciones a corto plazo entre mayor sea es mejor, porque evidencia que puede respaldar las deudas. (Entre mayor sea el indicador es mejor).

4. Puntaje VRC es un valor numérico de 150 a 950 y las entidades usan este puntaje para determinar el nivel de riesgo o incumplimiento de pago, entre más alto sea el puntaje menor es el riesgo y tendrá mayor posibilidad a acceder a un crédito para el análisis el promedio de puntaje VRC es de 685.

Tabla No8 (Fuente Propia)

Variables	Centroides	
	clúster 1	clúster 2
IND Apalancamiento	0,14	0,52
IND Endeudamiento	0,12	0,33
Razón Corriente	7,59	3,11
Puntaje VRC	685	582

El perfil de segmento óptimo para otorgamiento de crédito de caña panelera será aquel que tenga:

- Indicador de apalancamiento inferior al 14%
- Indicador de endeudamiento inferior al 12%
- Razón corriente por encima de 7.59
- Puntaje VRC superior a 685.

## Conclusiones

1. Para mejores decisiones en el otorgamiento de crédito se contribuyó en el desarrollo de la segmentación bajo el comportamiento de los clientes de caña panelera de una entidad bancaria.
2. Los segmentos de clientes a través de sus indicadores financieros están determinados por 2 clúster.
  - Grupo uno: clientes excelentes para cupo de crédito y desembolso en 5 días hábiles.
  - Grupo dos: clientes Buenos que ingresarían a evaluar en el proceso normal con los tiempos actualmente definidos de 10 a 15 días hábiles.
3. Los clientes grupo uno se caracterizan por:
  - Indicador de apalancamiento e indicador de endeudamiento bajos.
  - Indicador de Razón corriente puntaje VRC superiores.
4. Se desarrolló la metodología de agrupamiento de clientes basada en la segmentación K-Means conjuntamente con el análisis de ACP, esta información se actualizará mensualmente para fines comerciales.

## **Recomendaciones**

1. Para medir la performance del modelo de segmentación de clientes, se recomienda realizar el seguimiento durante los seis meses posteriores.
2. El modelo de segmentación está dirigido a clientes con suficiente información financiera, para aquellos que tienen poca se podría desarrollar un modelo de perfil con variables globales no necesariamente financieras (sociodemográficas).
3. Se realizó metodología con Análisis de Componentes principales y K-Means, para posteriores estudios se recomienda mirar clúster de Bietapico o arboles de decisión y potencializar el modelo definido.
4. Aplicar este estudio para todos los productos del Banco e incluir más dimensiones con el fin de ser más competitivos en el mercado.

## Bibliografía

- Andres Murcia Pabón (2007). Determinantes del acceso al crédito de los hogares colombianos. Borradores de economía, 449, 1-36.
- Cleger-Tamayo, S., Picanço-Bastos, M. S., & Figueiras-de-Farias, M. S. (2017). Evaluación de interfaces de explicación en los sistemas de recomendación. Revista científica, (29), 129-139.
- Demey, J., Adams, M. y Freites, H. (1992). Uso Del Método De Análisis De Componentes Principales Para La Caracterización De Fincas Agropecuarias. Agronomía Tropical, 44, 475-497.
- Forero, S. L., Ballesteros, V. B., & Toledo, J. L. N. (2018). Improved Likelihood Ratio Tests in Power Series Generalized Nonlinear Models. International Journal of Applied Engineering Research, 13(22), 15798-15805.
- Forero, SL, Ballesteros, VB y Toledo, JLN (2018). Estadística de gradiente: una opción para realizar pruebas de hipótesis en escenarios de tamaño de muestra pequeño. Revista Internacional de Investigación de Ingeniería Aplicada, 13 (23), 16368-16375.
- Gutiérrez, J., Castaño, N. y Asprilla, E. (2014). Mecanismos de financiación y gestión de recursos financieros del sector agropecuario en Colombia. Cenes Vol. 33, 147-176.
- Held, G. (1999). Políticas de crédito para empresas de menor tamaño con bancos de segundo piso: experiencias recientes en Chile, Colombia y Costa Rica. CEPAL, 84, 1-44.
- Hernández, C. (2014). Estrategias Para Aumentar Las Utilidades De La Producción Panelera En Colombia. Una descripción para aumento de utilidades en el producto de panela en Colombia. Bogotá: Universidad Militar Nueva Granada.
- Mestries, F. (2000). Globalización, crisis azucarera y luchas cañeras en los años noventa. Sociológica, 15, 41-68.
- Monroy-García, O. Y., Rolón-Quintero, D. C., & Sepúlveda-Mora, S. B. (2017). Sistema inalámbrico modular de bajo costo para supervisar variables ambientales

- en invernaderos: un respaldo al agricultor de pequeña escala. *Revista científica*, 2(29), 164-179.
- Moreno-Torres, O. H., Ballesteros-Ricaurte, J. A., & González-Sanabria, J. S. (2015). Transformación De Archivos Dicom A Formatos XML, JPEG y PNG- Transformation of Dicom Files to XML, JPEG and PNG Formats. *Revista científica*, 1(21), 71-80.
- Ochoa, J., Galeano, W. y Agudelo, L. (2010). Construcción de un modelo de scoring para el otorgamiento de crédito en una entidad financiera. *Perfil de Coyuntura Económica*, 16, 191-222.
- Orozco, A., Llano, O. y Ramírez, G. (2016). Sistemas de información enfocados en tecnologías de agricultura de precisión y aplicables a la caña de azúcar. *Revista Ingenierías Universidad de Medellín*, vol. 15, núm. 28, 103-124.
- Programa De Productividad Y Competitividad Agropecuaria Del Huila (2010). Tecnología de punta para el sector panelero, un compromiso institucional. Huila: Rivera, J.
- Rebollo, M. (2002). Aplicación del Análisis de Correspondencias Múltiples. *Revista de Investigación Educativa*, 20, 131-151.
- Romel y Acevedo. (2001). Los bancos de desarrollo Agrícola y el acceso al crédito rural. Ponencia presentada en la Reunión del Part-nership for Capacity Building in Rural Finance (CABFIN). 6 y 7 de diciembre. FAO, Roma, Italia.
- Tello, M., Eslava, H. y Tobias, L. (2013). Análisis y evaluación del nivel de riesgo en el otorgamiento de créditos financieros utilizando técnicas de minería de datos. *Vision Investigadora*, 1, 13-26.
- Trejos-Angel, D. Y., & Ramírez-Céspedes, O. A. (2014). Modelo matemático y simulaciones numéricas para un problema de frontera libre ecológico- Mathematical model and numerical simulations for a free boundary problem of ecological. *Revista científica*, 3(20), 128-137.
- Villaroel, L., Alvarez, J. y Maldonado, D. (2003). Aplicación del Análisis de Componentes Principales en el Desarrollo de Productos. Una descripción de

aplicación de componentes principales en un problema de desarrollo de un producto. Cochabamba: Universidad Mayor de San Simon.

Espinoza Tovar, A. S., & Espinoza Tovar, A. J. (2014). Diseño e implementación de políticas de crédito y cobro para reducir el riesgo de liquidez en la empresa Republicneg SA, ubicada en el cantón Samborondón, provincia del Guayas (Bachelor's thesis).

Barajas, A., López-Enciso, E. A., & Oliveros, H. (2001). ¿Por qué en Colombia el crédito al sector privado es tan reducido? Borradores de Economía; No. 185.

Gavilán, J., & Gallego -Torres, A. P. (2016). Implementación del modelo Lean Service en el proceso de recaudo de la Cooperativa de Ahorro y Crédito Fincomercio Ltda. Redes De Ingeniería, 7(2), 138-147.

Gallego-Torres, A. P., & Millán-Rojas, E. E. (2016). Innovación tecnológica para el desarrollo sustentable-Technological innovation for sustainable development. Revista científica, 3(26), 1-3.

Acevedo y Delgado. (2002, Septiembre). Desarrollo de las Economías Rurales en América Latina y el Caribe: Manejo Sostenible de los Recursos Naturales, Acceso a Tierras y Finanzas Rurales. Presentado al Banco Interamericano de desarrollo, Departamento de desarrollo sostenible, Fortaleza, Brasil

## Apéndice

### Código en R para ACP

```
install.packages("agricolae")
install.packages("ggplot2")
install.packages("readxl")
install.packages("FactoMineR")
install.packages("factoextra")
install.packages("MASS")
install.packages("ade4")
install.packages("ca")
install.packages("var")
install.packages("gg")
install.packages("paqueteadp")
```

```
library(agricolae)
library(ggplot2)
library(readxl)
library(FactoMineR)
library(factoextra)
library(MASS)
library(ade4)
library(ca)
library(var)
library(paqueteadp)
```

```
setwd("D:/Backup Toshiba/JESSI/Especialización_estadistica/Proyecto I/Proyecto II/A.modelo
SPSS y R proyecto Grado")
base_cana <- read.table("BASE_FINAL_CANA_PANELERA_VAR4.txt", header = T, sep =
"\t")
View(base_cana)
names(base_cana)
data_cont <-
base_cana[,c("IND_crecimiento_ventas", "IND_Apalancamiento", "IND_Endeudamiento", "Cobe
rtura_Interes", "Razon_Corriente", "Rentabilidad_Sobre_Activos", "Rentabilidad_Sobre_Patrimo
nio", "IND_Desarrollo_Nuevos_Productos", "puntaje_VRC")]
base_PCA <- data.frame(data_cont)
names(base_PCA)
base_PCA <- data.frame(data_cont)
cat("\n Matriz de correlaciones \n")
print(cor(base_PCA), 2)
plot(base_PCA)
acp <- dudi.pca(base_PCA, scannf=F, nf=7)
acpI <- inertia.dudi(acp, row.inertia=T, col.inertia=T)
```



```

cat("\n Valores propios \n")
print(acpI$TOT,2)
plot(acp$eig)
print(acp$cl)

cat("\n Coordenadas de las columnas \n")
print(acp$co)
cat("\n Contribuciones de las columnas a los ejes \n")
print(acpI$col.abs/100)
cat("\n Calidad de representaci3n de las columnas \n")
print(acpI$col.rel/100)
cat("\n Calidad de representaci3n de las columnas en el plano \n")
print(acpI$col.cum/100)
cat("\n Coordenadas de las filas \n")
print(acp$li)
cat("\n Contribuciones de las filas a los ejes \n")
print(acpI$row.abs/100)
cat("\n Calidad de representaci3n de las filas en los ejes \n")
print(acpI$row.rel/100)
cat("\n Calidad de representaci3n de las filas en el plano \n")
print(acpI$row.cum/100)
par(mfrow=c(1,1)) # para 4 gráficas simultáneas
s.corcircle(acp$co,sub="Variables - Círculo de correlaciones",possub= "bottomright")
datosE=scale(base_PCA)
acp = prcomp(datosE)
summary(acp)

```

### Segundo código

```

datosE1=scale(data_cont1)
cor(datosE1)
cov(datosE1)
acp = prcomp(datosE1)
summary(acp)
res.pca <-PCA(data_cont,graph =FALSE)
print(res.pca)
# Obtención de los eigenvalores
eig.val <-eig.val(res.pca)
eig.val
fviz_eig(res.pca,addlabels =TRUE,ylim =c(0,80))
var$coord
head(var$cos2)
head(var$contrib)
#Círculo de correlación#
head(var$coord,7)
fviz_pca_var(res.pca,col.var ="black")
res.pca
summary(res.pca)

```